

TSM

Modelování molekulárních struktur

Cvičení III

Petr Kulhánek

petr.kulhanek@ceitec.muni.cz

CEITEC – Středoevropský technologický institut, Masarykova univerzita,
Kamenice 5, 625 00 Brno

NCBR – Národní centrum pro výzkum biomolekul, Masarykova univerzita,
Kotlářská 2, 611 37 Brno

Dávkové systémy

Dávkové zpracování

Dávkové zpracování je vykonávání série programů (tzv. dávek) na počítači bez účasti uživatele. Dávky jsou připraveny předem, takže mohou být zpracovány předány bez účasti uživatele. Všechna vstupní data jsou předem připravena v souborech (skriptech) nebo zadána pomocí parametrů na příkazovém řádku. Dávkové zpracování je opakem interaktivního zpracování, kdy uživatel až teprve za běhu programu poskytuje požadované vstupy.

Výhody dávkového zpracování

- sdílení zdrojů počítače mezi mnoha uživateli a programy
- odložení zpracování dávek do doby, kdy je počítač méně vytížen
- odstranění prodlev způsobeným čekáním na vstup od uživatele
- maximalizace využití počítače zlepšuje využití investic (zejména u dražších počítačů)

zdroj: www.wikipedia.cz, upraveno

Nástroje pro dávkové spouštění

➤ **OpenPBS**

<http://www.mcs.anl.gov/research/projects/openpbs/>

➤ **PBSPro**

<http://www.pbsworks.com>

➤ **Oracle Grid Engine**

<http://www.oracle.com/us/products/tools/oracle-grid-engine-075549.html>

➤ **Open Grid Scheduler**

<http://gridscheduler.sourceforge.net/>

open source

➤ **Torque**

<http://www.adaptivecomputing.com/products/open-source/torque/>

je použit jako dávkový systém v MetaCentrum VO, na klastrech SOKAR a WOLF

Infinity

<https://lcc.ncbr.muni.cz/whitezone/development/infinity/>

Přehled příkazů

Správa software:

- site aktivace logických výpočetních zdrojů
- module aktivace/deaktivace software

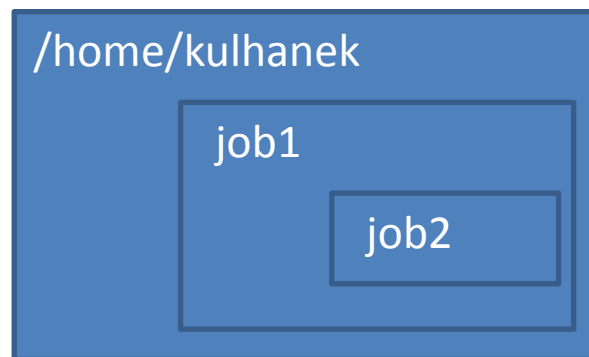
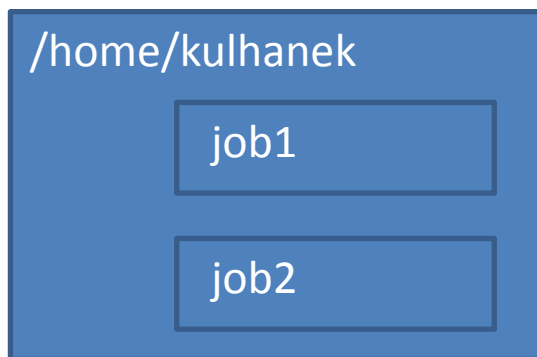
Správa úloh:

- pqueues přehled front z dávkového systému dostupných uživateli
- pnodes přehled výpočetních uzlů dostupných uživateli
- pqstat přehled všech úloh zadaných do dávkového systému
- pjobs přehled úloh uživatele zadaných do dávkového systému
- psubmit zadání úlohy do dávkového systému
- pinfo informace o úloze
- pgo přihlásí uživatele na výpočetní uzel, kde se úloha vykonává
- paliases definování aliasů

Úloha

Úloha **musí splňovat** následující podmínky:

- každá úloha se spouští v samostatném adresáři
- všechny vstupní data úlohy musí být v adresáři úlohy
- adresáře úloh nesmí být do sebe zanořené
- průběh úlohy je řízen skriptem nebo vstupním souborem (u automaticky detekovaných úloh)
- skript úlohy musí být v bashi
- ve skriptu úlohy se nesmí používat absolutní cesty, všechny cesty musí být uvedeny relativně k adresáři úlohy



Spuštění úlohy

Úlohu spouštíme **v adresáři úlohy** příkazem **psubmit**.

```
psubmit destination job [resources] [syncmode]
```

destination (kam) je buď:

- název_fronty
- název_uzlu@název_fronty

job je buď:

- název skriptu úlohy
- název vstupního souboru pro automaticky rozpoznávané úlohy

resources jsou požadované zdroje pro úlohu, pokud není uvedeno, požaduje se běh na 1 CPU

syncmode určuje způsob kopírování dat mezi adresářem úlohy a výpočetním uzlem, výchozím módem je "sync"

Monitorování běhu úlohy

K monitorování průběhu úlohy lze použít příkaz **pinfo**, který se spouští buď v adresáři úlohy nebo v pracovním adresáři na výpočetním uzlu. Dalšími možnostmi jsou příkazy **pjobs** a **pqstat**.

Pokud je úloha spuštěna na výpočetním uzlu, je možné použít příkaz **pgo**, který se naloguje na výpočetní uzel a změní aktuální adresář do pracovního adresáře úlohy.

Servisní soubory

V adresáři úlohy vznikají při zadání úlohy do dávkového systému a dále v průběhu života úlohy a po jejím ukončení servisní soubory. Jejich význam je následující:

- *.info kontrolní soubor s informacemi o průběhu úlohy
- *.infex vlastní skript (wrapper), který se spouští dávkovým systémem
- *.infout standardní výstup z běhu *.infex skriptu, **nutno analyzovat při nestandardním ukončení úlohy**
- *.nodes seznam uzlů vyhrazených pro úlohu
- *.gpus seznam GPU karet vyhrazených pro úlohu
- *.key unikátní identifikátor úlohy
- *.stdout **standardní výstup z běhu skriptu úlohy**

Cvičení I

1. Vytvořte textový soubor obsahující následující text. Soubor nazvěte **get_hostname** a uložte jej do samostatného adresáře **uloha01**.

```
#!/bin/bash
hostname
```

2. Nastavte souboru práva pro spuštění:

```
$ chmod u+x get_hostname
```

3. Vytvořený skript spusťte v terminálu. Co se stane?

```
$ ./get_hostname
```

4. Skript zadejte ke spuštění do dávkového systém. K zadání úlohy použijte frontu **short**.

```
$ psubmit short get_hostname
```

5. Průběh úlohy monitorujte příkazem **pinfo**.

6. Na kterém výpočetním uzlu se úloha spustila? Analyzujte obsah souboru **get_hostname.stdout** v textovém editoru.

Cvičení II

1. Vytvořte textový soubor obsahující následující text. Soubor nazvěte **long_job** a uložte jej do samostatného adresáře **uloha02**. Co dělá příkaz **sleep** (použijte manuálové stránky)?

```
#!/bin/bash
hostname
sleep 240
```

2. Skript zadejte ke spuštění do dávkového systém. K zadání úlohy použijte frontu **short**.

```
$ psubmit short long_job
```

3. Průběh úlohy monitorujte příkazem **pinfo**. Úlohy ostatních uživatelů příkazem **pqstat**. Obsazení jednotlivých výpočetních uzlů pak příkazem **pnodes**.

Výpočty v programu gaussian

Gaussian

<http://www.gaussian.com/>

Komerční program určený převážně pro kvantově chemické výpočty. Instalovaný na klastru WOLF.

Spouštění výpočtu

Výpočty v programu gaussian provádíme na **výpočetním klastru WOLF**.

1) Aktivace modulu gaussian:

pouze jednou v daném terminálu

```
$ module add gaussian:03.E1
```

2) Přímé spuštění výpočtu (**POUŽÍVAT JEN PRO KRATKÉ VÝPOČTY**):

```
$ g03 soubor
```

jméno vstupního souboru bez přípony .com

Po skončení výpočtu bude v adresáři nový soubor (**soubor.log**) obsahující výsledky výpočtu.

Spouštění výpočtu, II

3) Dávkové spouštění (**PREFEROVANÝ ZPŮSOB**):

```
$ psubmit localhost@short soubor.com
```

výpočet spustí na **lokálním** stroji ve frontě **short**

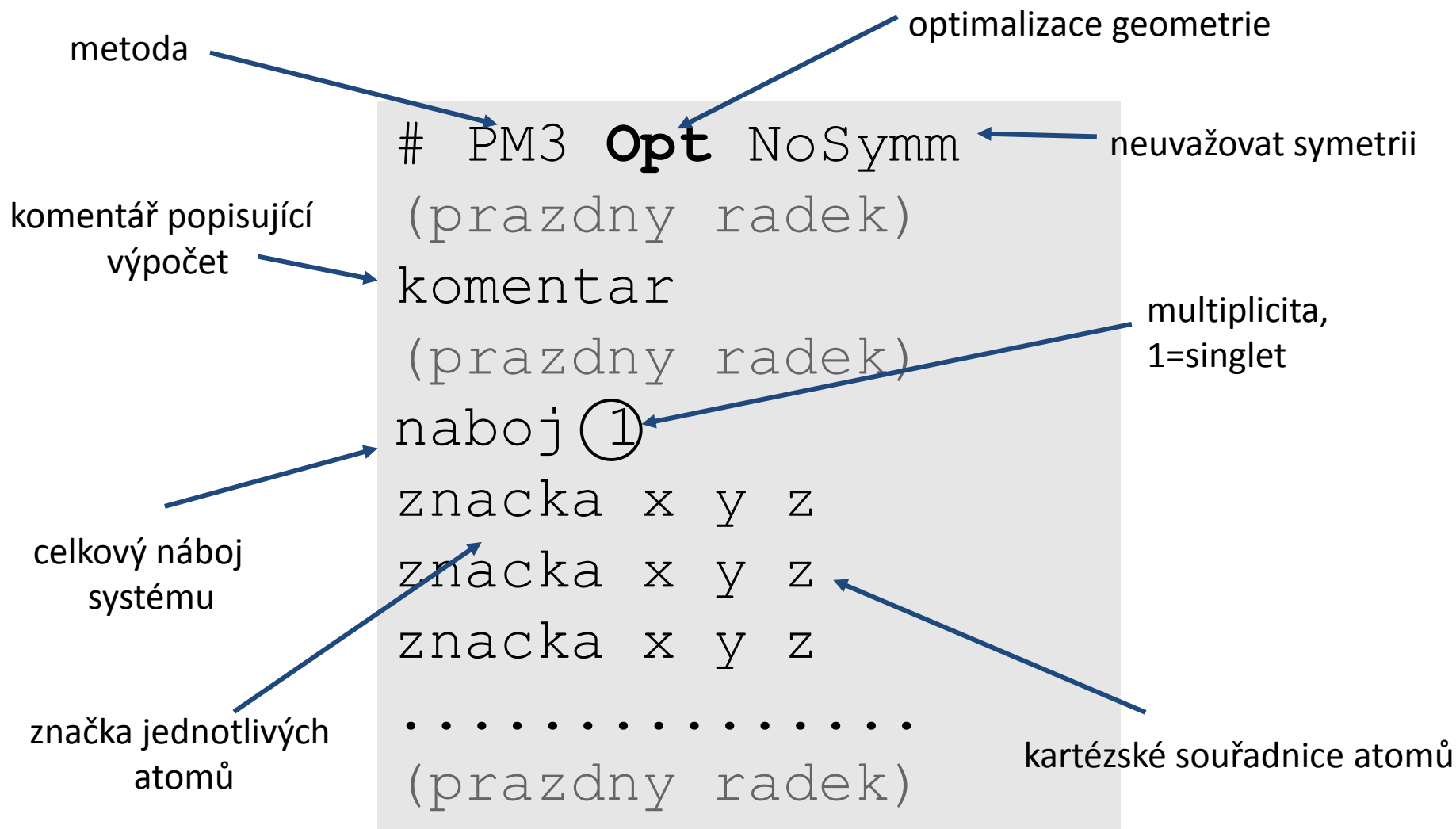
jméno vstupního souboru **včetně** přípony .com
skript úlohy se vytvoří automaticky

Nezapomeň:

- každý výpočet musí být spuštěn v samostatném adresáři!!

Optimalizace geometrie

Optimalizace geometrie, vstup



soubor ukládáme s příponou **.com**

Opt. geometrie, ukázka vstupu

ethan.com

kartézské souřadnice atomů v Ångströmech

```
# PM3 Opt Nosymm
```

```
Ethan
```

```
0 1
```

C	-8.17932	3.14871	-0.03132
C	-7.00267	2.20120	0.03204
H	-7.90498	4.06824	-0.55607
H	-8.51264	3.41431	0.97740
H	-9.01757	2.68656	-0.56001
H	-6.16234	2.66559	0.55546
H	-6.67317	1.93085	-0.97667
H	-7.27527	1.28418	0.56205

prázdný řádek

Příprava vstupního souboru, NEMESIS

- 1) Projekt: Build Structure
- 2) Sestavíme molekulu
- 3) Provedeme optimalizaci pomocí molekulové mechaniky
- 4) File->Export Structure as -> Gaussian Input
- 5) Nastavíme metodu a typ výpočtu, uložíme

Project 1: NEMESIS - Molecular Modelling Package

Structures

Name	SID	Ato
Structure 1	1	

Build panel

Basic General

Optimize

Geometry

Position Distance Angle Torsion

Restrain Property Label

Gaussian Input

Title: Title

Calculation: Geometry Optimization Processors: 1

Theory: PM3 Basis:

Charge: 0 Multiplicity: 1

Format: Cartesian

Checkpoint Include PBC Include fragments

```
%Chk=checkpoint.chk
# RPM3 Opt NoSymm
```

Title

O	1			
H	0.043916	-0.336071	1.404725	
C	-0.017513	-0.336324	0.312494	
H	-1.075148	-0.371158	0.031690	
H	0.453524	-1.259117	-0.043340	
C	0.702669	0.874594	-0.267004	
H	1.562094	0.531237	-0.853870	

Save setup Close on save

Reset Save ... Close

S T C G D P select a connect atom (open valence) to attach a molecular fragment

Optimalizace geometrie

PM3

Výsledný soubor

1) Otevřete výstupní soubor v textovém editoru:

```
$ kwrite soubor.log &> /dev/null &
```

2) Projděte souborem, pouvažujte nad významem jednotlivých sekcí:

Input orientation:

```
.....  
SCF Done: E(RPM3) = -0.289342988150E-01 A.U. after 8 cycles  
.....  
Center Atomic Forces (Hartrees/Bohr)  
.....  
Item Value Threshold Converged?  
Maximum Force 0.000311 0.000450 YES  
RMS Force 0.000068 0.000300 YES  
Maximum Displacement 0.009047 0.001800 NO  
RMS Displacement 0.004622 0.001200 NO  
.....  
Optimization completed.  
-- Stationary point found.  
.....  
Normal termination of Gaussian 09
```

Načtení výstupního souboru, NEMESIS

- 1) Projekt: Trajectory
- 2) File->Import Trajectory as -> Gaussian Geometry Optimization

The screenshot displays the NEMESIS Molecular Modelling Package interface. The main window shows a 3D ball-and-stick model of an ethane molecule. A blue arrow labeled "dvojklik" (double-click) points from the "Trajectory 1 Structure 1" entry in the "Trajectories" panel to the "Trajectory" dialog box. This dialog box has tabs for "Basic", "Play", "Segments", "Filters", and "Referenced by". Another blue arrow labeled "dvojklik" points from the "Trajectory" dialog to the "Gaussian Geometry Optimization" dialog box. The "Gaussian Geometry Optimization" dialog has tabs for "Basic", "Energy", and "Info". The "Energy" tab is active, showing a table of energy values for six snapshots.

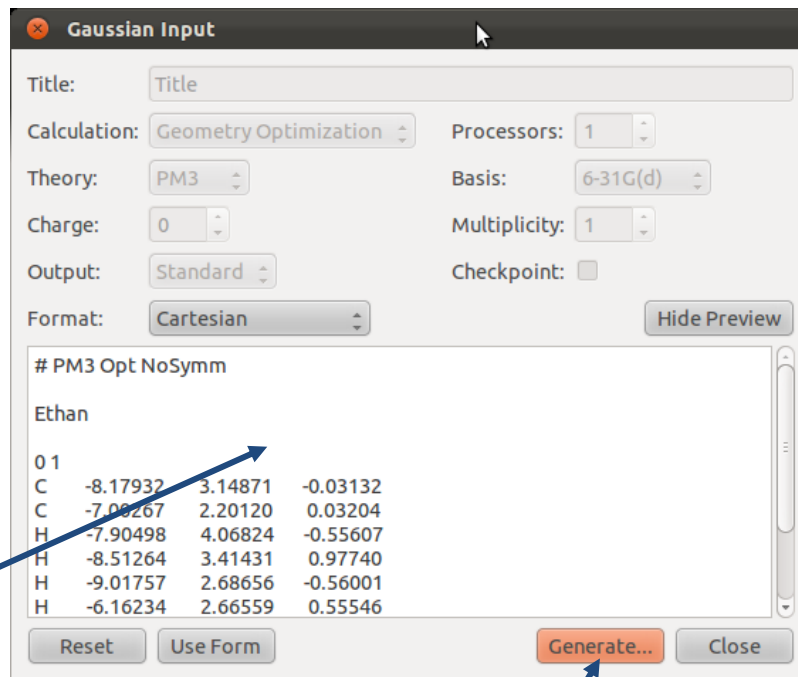
ID	Energy [a.u.]	Relative Energy [kcal]
1	-0.028617175	0.00
2	-0.028893932	-0.17
3	-0.028929498	-0.20
4	-0.028931879	-0.20
5	-0.028934299	-0.20
6	-0.028935040	-0.20

At the bottom of the interface, a blue oval highlights a set of playback controls (stop, play, next, previous, etc.), with a blue arrow pointing to it from the text "průběh optimalizace" (optimization process).

Optimalizace geometrie alternativní postupy

Příprava vstupního souboru, Avogadro

Vstupní soubor pro výpočet můžeme vytvořit pomocí programu **Avogadro**. Příslušný nástroj naleznete v menu **Extensions->Gaussian**.



ručně upravíme
podle typu výpočtu

vytvoří vstupní soubor pro výpočet

Pokud je struktura výsledkem **předchozího výpočtu** v programu gaussian, **neměli** bychom tuto strukturu v programu Avogadro **měnit** (např. optimalizovat geometrii)!

Optimalizace geometrie, výsledek

1) Aktivace modulu qmutil:

```
$ module add qmutil
```

pouze jednou v daném terminálu

2) Zobrazení průběhu optimalizace (energie):

```
$ extract-gopt-ene soubor.log
```

3) Průběh optimalizace (všechny geometrie):

```
$ extract-gopt-xyz soubor.log > soubor_opt.xyz
```

4) Získání optimalizované geometrie (poslední):

```
$ extract-xyz-str soubor_opt.xyz last > soubor_last.xyz
```

Je vhodné podívat se na průběh optimalizace, např. v programu **vmd**, **Avogadro**, nebo **Nemesis**.

Optimalizace geometrie, výsledek

Průběh optimalizace (energie):

```
[kulhanek@wolf ethan]$ extract-gopt-ene soubor.log
```

```
# Coordinate:
```

# Step	Energy [kcal/mol]	Energy [au]
1	0.000	-0.028650961
2	-0.171	-0.028922822
3	-0.188	-0.028950914
4	-0.190	-0.028953934

číslo optimalizačního kroku

relativní energie vůči **výchozí geometrii**

absolutní energie v Hartree

Energie optimalizované struktury, tedy geometrie obsažené v **soubor_last.xyz**, v Hartree.

Vibrační analýza

Výsledkem vibrační analýzy jsou **normální vibrace** (lineárně nezávislé pohyby atomů v molekule) odpovídající spřaženým **harmonickým oscilátorům** (jedná se o aproximaci). Vibrační analýzu lze využít k předpovědi absorpčních spekter v oblasti infračerveného záření (není příliš přesné) nebo k **určení typu stacionárního bodu na ploše potenciální energie**. Lze tak od sebe odlišit **optimální strukturu** (všechny frekvence vibrací jsou **kladné**) od **tranzitního stavu** (**právě jedna** frekvence musí být **imaginární**).

Vibrační analýza se provádí pro geometrii, která byla

optimalizována stejnou metodou,

jaká bude použita pro vibrační analýzu. V opačném případě budou spočtené frekvence zcela chybné.

Vibrační analýza, vstup

frekvenční analýza

```
# PM3 FREQ NoSymm
(prazdny radek)
komentar
(prazdny radek)
naboj 1
znacka x y z
znacka x y z
znacka x y z
.....
(prazdny radek)
```

optimalizovaná
geometrie

soubor ukládáme s příponou **.com**

Příprava vstupního souboru, NEMESIS

- 1) **Optimalizovaná** geometrie
- 2) File->Export Structure as -> Gaussian Input
- 3) Nastavíme metodu a typ výpočtu, uložíme

Project 1: NEMESIS - Molecular Modelling Package

Calculation: **Frequencies**

Theory: **PM3**

Format: Cartesian

Atom	X [a.u.]	Y [a.u.]	Z [a.u.]
O 1			
H	-0.908150	0.201192	0.637090
C	-0.034170	-0.058327	0.025579
H	-0.390627	-0.686639	-0.801043
H	0.623995	-0.680150	0.646308
C	0.675174	1.169534	-0.477182
H	1.549213	0.910014	-1.088609

Energy [a.u.] Relative Energy [kcal/mol]

Energy [a.u.]	Relative Energy [kcal/mol]
-0.028617175	0.00
-0.028893932	-0.17
-0.028929498	-0.20
-0.028931879	-0.20
-0.028934299	-0.20
-0.028935040	-0.20

Poslední (optimalizovaná) geometrie

Frequencies

PM3

Výsledný soubor

1) Otevřete výstupní soubor v textovém editoru:

```
$ kwrite soubor.log &> /dev/null &
```

2) Projděte souborem, pouvažujte nad významem jednotlivých sekcí:

Input orientation:

```
....  
SCF Done: E(RPM3) = -0.289350395641E-01 A.U. after 9 cycles....  
....  
Harmonic frequencies (cm**-1), IR intensities (KM/Mole), Raman scattering  
activities (A**4/AMU), depolarization ratios for plane and unpolarized  
incident light, reduced masses (AMU), force constants (mDyne/A),  
and normal coordinates:  
          1              2              3  
          A              A              A  
Frequencies -- 224.6457          878.1828          878.1936....
```

```
.....  
-----  
- Thermochemistry -  
-----
```

http://gaussian.com/g_whitepap/thermo.htm

```
...  
Zero-point vibrational energy      194420.2 (Joules/Mol)  
...  
Item          Value      Threshold  Converged?  
Maximum Force      0.000005    0.000450    YES  
RMS Force          0.000002    0.000300    YES  
Maximum Displacement 0.000091    0.001800    YES  
RMS Displacement  0.000049    0.001200    YES  
...  
Normal termination of Gaussian 09 at Sun Mar 10 20:23:26 2013.
```

Vibrační analýzu lze provádět pouze na stacionárních bodech (optimalizovaných geometriích).

Načtení výstupního souboru, NEMESIS

- 1) Projekt: Trajectory
- 2) File->Import Trajectory as -> Gaussian Vibrations

The screenshot shows the NEMESIS Molecular Modelling Package interface. The main window displays a ball-and-stick model of a molecule. A blue arrow labeled "dvojklik" (double click) points from the "Trajectory 1 Structure 1" entry in the "Trajectories" panel to the "Trajectory" dialog box. The "Trajectory" dialog box has a table with the following data:

SI	Name	Snapshots	Type
1	ethan_freq	180	Gaussian Vibratic

Another blue arrow labeled "dvojklik" points from the "Trajectory" dialog box to the "Gaussian Vibrations" dialog box. The "Gaussian Vibrations" dialog box has a table with the following data:

ID	Frequency	IR Intensity	Scale
1	224.6	0.0	
2	878.2	0.0	
3	878.2	0.0	
4	1120.0	0.0	
5	1120.0	0.0	
6	1137.8	0.0	
7	1359.3	0.0	
8	1408.2	0.0	
9	1408.2	0.0	
10	1443.6	0.0	

A blue arrow labeled "zvolíme vibraci" (we choose vibrations) points to the checkboxes in the "Gaussian Vibrations" dialog box. Below the dialog box, a blue arrow labeled "spustíme animaci" (we start animation) points to the animation control buttons at the bottom of the interface.

Vibrační analýza alternativní postupy

Vibrační analýza, výstup - Avogadro

Do programu Avogadro načteme **soubor.log**, obsahující výsledky vibrační analýzy. Souhrn frekvencí jednotlivých normálních vibrací najdeme v menu **Extensions->Vibrations**.

The screenshot shows the Avogadro interface with the 'Molecular Vibrations' dialog box open. The dialog displays a table of vibrational frequencies and intensities. A 3D ball-and-stick model of a molecule is visible in the background. Two blue arrows point from the dialog to labels: one from the 'Start Animation' button to 'vizualizace vibrací' and another from the table to 'frekvence vibrací'.

	Frequency (cm ⁻¹)	Intensity (km/mol)
1	223.7	0.0
2	877.8	1.5
3	878.0	1.5
4	1,119.8	0.0
5	1,119.9	0.0
6	1,138.0	0.0
7	1,359.2	0.3
8	1,408.1	0.3

Options:

- Scale: [Slider]
- Display force vectors
- Animation speed set by frequency

Buttons: Start Animation, Export..., Close

Vibrační analýza, výstup - Molekel

Molekel

<http://molekel.cscs.ch/wiki/pmwiki.php>

Program pro vizualizaci molekulárních struktur.

1) Aktivace modulu molekel:

```
$ module add molekel
```

2) Otevření programu molekel:

```
$ molekel
```

3) Načíst do programu soubor *soubor.log*:

4) Animation->Per molecule settings ...

.Animation (Tab) → Animation mode → Vibration

.Vibration (Tab) → zvolit danou vibraci

4) Animation->Start animation

<http://www.youtube.com/watch?v=Aan97dvvVqk>

Cvičení

1. Namodelujte molekulu formaldehydu. Najděte optimální geometrii. K výpočtu energie použijte semiempirickou kvantově chemickou metodu PM3. Jaká je délka vazby C=O?
2. Proveďte vibrační analýzu na optimalizované geometrii. Jaká je charakteristická vibrace vazby C=O?